DOCUMENT RESUME

ED 060 546 EA 004 114

TITLE	An Experiment in Performance Contracting. Summary of Preliminary Results.
INSTITUTION	Office of Economic Opportunity, Washington, D.C. Office of Planning, Research, and Evaluation.
REPORT NO	OEO-3400-5
PUB DATE	1 Feb 72
NOTE	40p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Academic Achievement; *Disadvantaged Youth; *Educational Experiments; Evaluation Methods; Federal Programs; Mathematics Education; *Performance Contracts; Reading Skills; Standardized Tests

ABSTRACT

This report describes briefly the experimental design and presents the basic contract provisions. The experiment results reveal that performance contracting is no more successful than traditional classroom methods in improving the reading and mathematics skills of poor children. Both control and experimental groups performed equally poorly in terms of overall averages. The report concludes that the evidence fails to indicate that performance contracting will bring about any great improvement in the educational status of disadvantaged children. (Author/JF)

ED 060546

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

AN EXPERIMENT

IN PERFORMANCE CONTRACTING

SUMMARY OF PRELIMINARY RESULTS

Office of Economic Opportunity

Office of Planning, Research, and Evaluation

February 1, 1972

EA 004 114

May 14, 1970

"If it turns out that there are elements of this that prove successful, one would think it would have the potential for affecting public policy with respect to education.

"If the results prove that all the approaches that we utilize within the umbrella of the total experiment are not successful and not desirable, the evaluation will indicate that. By the same token, the experiment still will affect policy because it will lead us to the conclusion that performance contracting is not a desirable route to go."

-- Donald Rumsfeld
Former Director
Office of Economic Opportunity

PREFACE

The information in this pamphlet is based on a preliminary analysis of the data from the Office of Economic Opportunity experiment in performance contracting in education. The issues summarized here will be discussed in more detail in another volume, OEO pamphlet 3400-6, which will be available about March 1, 1972. It will include:

-- A more technical and comprehensive analysis of the aggregate evaluation test results.

-- A description of the standardized tests that were used for the evaluation, and the issues surrounding their relevance for a project of this nature.

-- A description of the contracts between the OEO and the school districts and between the school districts and the private technology firms, of the incentives structure used to determine the firms' payments, and of problems that arose in the implementation of the contracts.

-- A statement from the local project directors on their perceptions of the experiment.

-- An analysis of the costs involved in implementing the performance contracts.

Another report will be issued in about 15 days on a related experiment in which teachers' groups, rather than private technology firms, contracted with their school districts to provide educational services on an incentives basis.

Additional information also will be available in the Interim Report on the OEO Experiment in Performance Contracting prepared by the Battelle Memorial Institute, the testing and analysis contractor for the experiment.* The OEO analysis summarized here emphasizes comparisons of aggregate results from the control and experimental groups; the Battelle interim report, in addition to providing a detailed description of the experiment's operation, emphasizes comparisons of the evaluation test results on a site by site basis. Finally, data tapes will be available at the cost of reproduction. These may be obtained from Charles Stalford, project manager for the experiment by interested researchers.

It should be emphasized that the results discussed in the two OEO volumes and the Battelle interim report are preliminary. The broad conclusions that are outlined here can be viewed with confidence, but idiosyncrasies concerning sample characteristics, testing conditions, and other factors necessitate that caution be used when results for individual sites are examined. Much further analysis is required

---

*Copies of the Battelle report will be available from the National Technical Information Service, U. S. Department of Commerce, Springfield, Virginia 22151. The final report of the management support contractor, Education Turnkey Systems, also is available from the Information Service. Entitled Final Report to the Office of Economic Opportunity: Performance Incentive Remedial Education Experiment PB 202830, its cost is $3.00. Another useful research reference is a Rand Corporation evaluation funded by the U. S. Department of Health, Education, and Welfare. The six-volume report, R-900/1-6-HEW, Case Studies in Educational Performance Contracting, includes Grand Rapids, one of the OEO's experiment sites, in its case studies.

before the site by site results can be fully understood or explained.

The OEO will continue its analysis in an attempt to further refine and extend the results summarized here. In addition, further analysis will be included in the final Battelle report, expected later this winter. That report also will include discussions of:

-- Retention tests administered at sites where there was some early indication that children in the experimental group improved at a significantly better of worse rate than children in the control group.

-- The results of a questionnaire filled out by parents of children in the experiment. The questionnaire concerned parents' attitudes toward education in general and the performance contracting experiment in particular.

-- Results of tests administered to children in the comparison and special treatment groups.

-- An analysis of the impact of performance contracting on absenteeism.

This experiment could not have been accomplished without the extraordinary assistance and cooperation of a number of individuals. Twenty-four hour days and seven-day work weeks were required of the management support contractor, Education Turnkey Systems, during the start-up phase, and similar round-the-clock sieges faced the evaluation contractor, Battelle Memorial Institute, during the testing and analysis periods. The project directors frequently were called upon for resourcefulness, patience, and dedication far beyond the normal range of human

capabilities. Principals of the schools in which the experiment took place suffered inconveniences and disruptions to the normal operations of their schools with commendable toleration, while the district superintendents and school board members assured the experiment's success with their constant support.

Much credit is due also to those within OEO who were responsible for the experiment's conception and implementation. John Oliver Wilson, former Director of the Office of Planning, Research, and Evaluation, supervised the experiment from the time of his staff's first visit to Texarkana until November 1971, and John Evans, former Director of the Evaluation Division, contributed greatly to the design phase.

The staff of the experiment was headed by Jeffry Schiller, Director of the Experimental Research Division; Charles Stalford, the project manager; and Judy Glotzer, the assistant project manager. Working with them were Ellen Murdoch and Ernest Palmer, and two dedicated secretaries, Helen Duran and Margaret Parker. The bulk of the in-house analysis was undertaken by Edward Gramlich and Irwin Garfinkel, with the assistance of Jane Lee, Gary Liberson, Fritz Scheuren, and Les Klein. Melinda Upp provided editorial services. Invaluable assistance also was rendered by the OEO's Procurement Office, headed by Ralph Howard, and his staff, Mike Burke, Jim Bacon, George Boxall, Fred Hanau, Norton Olshin, and Rosemarie Lesineur. And, frequent support was provided

by the Office of the General Counsel and its staff, including Robert Trachtenberg, Paul Stone, Lawrence Weiner, and John Siegmund.

Thomas K. Glennan, Jr.
Acting Director
Office of Planning,
Research and Evaluation

TABLE OF CONTENTS

	Page
Preface	i
Introduction	1
The Experimental Design	5
Results	17
Contractual Procedures	25
Summary and Conclusions	31

TABLES

Table I:	Comparison of Particular Aspects of Experimental Programs	10
Table II:	Characteristics of Students	12
Table III:	Mean Gains of Experimental and Control Students Across All Sites	18
Table IV:	Status of Experimental Students Before and After Performance Contracting	20
Table V:	Evaluation Test Results (GEQ)	21
Table VI:	Summary of Significant Results at Individual Sites	23
Table VII:	Summary of Contractor Incentives	28

I.

INTRODUCTION

In many ways, public school education is better today than it has been at any time in our history: we are spending more money per pupil than ever before; children are learning more and learning it earlier; and illiteracy rates are dropping while the average years of schooling completed by our adult population is steadily increasing.

At the same time, general dissatisfaction with the public schools is increasing among taxpayers, who are turning down bond and tax rate increase referenda in larger proportions; among parents, who are demanding accountability and community control over schools; among educators, who have seen the failure of most current compensatory programs;[1/] and among legislators, who question whether the billions of dollars they have appropriated for public education have been wisely used.

These concerns are most acute among the poor, who correctly perceive the public education system as one of the most important--if not the only--route to eventual economic self-sufficiency for their children. While it is impossible to isolate all the factors contributing to the problem, it is clear that by almost any criterion, poor children are not succeeding in our public schools.

Thus, great enthusiasm and optimism greeted reports that a new program, called performance contracting, was succeeding beyond anyone's

---

1/ A recent survey of evaluations by the U.S. Office of Education found that 10 of the 1,200 compensatory programs that were evaluated were successful.

wildest hopes with poor children in Texarkana, Arkansas, and Liberty-Eylau, Texas. Initial indications were that the project was doubling--in some cases even tripling--previous achievement gains of poor children, that drop-out rates had declined dramatically, and that school vandalism had been nearly eliminated. Performance contracting emphasized not inputs (teacher-pupil ratios, dollar-per-pupil expenditures, etc.) but outputs, what the children actually learned. The performance contracting system was new to education, although it had been tried in other fields. Its elements are relatively simple:

-- A contractor signs an agreement to improve students' performance in certain basic skills by set amounts.
-- The contractor is paid according to his success in bringing students' performance up to those prespecified levels. If he succeeds, he makes a profit. If he fails, he doesn't get paid.
-- Within guidelines established by the school board, the contractor is free to use whatever instructional equipment, techniques, or incentive systems that he feels will work.

The Texarkana project, funded under Title VIII of the Elementary and Secondary Education Act, was intended primarily as a drop-out prevention program. It featured a heavy reliance on individualized instruction and on various audio-visual teaching aids, ingredients that were not in themselves particularly new or revolutionary. What

was unusual about the Texarkana project was the contractual arrangement between the school district and the private firm providing the instruction: The firm would be paid only to the extent that it improved the students' scores on standardized reading and math tests. If the students did not improve, the contractor would not be paid even for the costs. The contractor, in turn, extended the concept of incentives and accountability to teachers and the students. Teachers' incentives included stock in the company; the children were offered a variety of rewards, ranging from trading stamps to free time for recreational activities.

As reports of the Texarkana experience circulated among educators, dozens of school districts began to consider performance contracting to meet their own needs. Staff from the Office of Economic Opportunity also visited Texarkana and were encouraged by the concept's potential to help poor children. But, they were also concerned that this single project was not designed to provide educators with the information they needed to decide whether performance contracting would meet their own school's needs. The Texarkana project was designed primarily to demonstrate that drop-outs could be reduced by improving classroom achievement. It was not an experiment with a rigorous evaluation structure. And, even had Texarkana had the most scientific and best-designed evaluation system possible, it still could not have indicated whether results achieved there were a fluke, whether they could be replicated elsewhere, whether the system was administratively

adaptable for other districts, or whether costs would be prohibitive. It was clear, then, that a broad, clearly defined, and carefully evaluated experiment was needed before performance contracting could be judged.

Thus, the Office of Economic Opportunity decided to mount a nationwide experiment to provide information that educators and school boards needed before deciding whether to enter into performance contracting.

Shortly after this decision was made, new reports from Texarkana seemed to justify the OEO's caution and graphically illustrate the need for better controls. It was reported that the contractor had provided teachers with some of the same materials--the same questions, in fact--that the children would face when being tested. The children had done well, it was charged, because they had been asked the same questions so many times they could not have failed to learn the answers. At this point, the Texarkana experience is still so confused that it is impossible to state with any certainty just how much "teaching to the tests" took place or how badly the test results were contaminated. What is known is that the Texarkana project was successful in reducing the drop-out rate. But it provided no reliable indication of what can or should be expected of performance contracting in terms of educational achievement. The OEO's experiment was designed to provide such an indication.

THE EXPERIMENTAL DESIGN

School districts traditionally are forced to rely on an informal grapevine for information about new educational techniques or instructional methods that are "successful." One year one district tries something called "new math," for example; the next year four publishing houses have issued "new math" curricula, and the year after that, dozens of districts across the country have installed "new math" programs. A schol board's decision to adopt a new program, meanwhile, can be based on little more than optimism that the first district's criteria of "success" were the same as its own or that the "success," however it may have been defined, that has been touted for the program can be replicated in a different setting. Seldom are new techniques subjected to any kind of rigorous evaluation; when evaluations are undertaken, seldom are they done in such a way that they generate information with broad applicability.

The OEO's experiment in performance contracting, then, represents the first attempt to submit an educational fad to any sort of controlled scientific evaluation that would have nationwide relevance. The goals of the experiment were straightforward: It would test the capabilities of education technology firms to improve the reading and math abilities of under-achieving youngsters in the context of a performance, or incentive based, contract. The experiment would last for one academic year. And, as stated in the request for proposals from the firms, "The purpose of this experiment is to evaluate the relative effectiveness of existing techniques, not to underwrite the development of new techniques."

So that the experiment would have the broad applicability that prior experiments had lacked, it was decided to include both the primary and secondary grades, a range of student populations that would approximately represent the poverty population, and a variety of instructional techniques. And, rather than a single observation, as was provided by Texarkana, the experiment would include a number of geographically dispersed school districts.

It was hoped that within this context, the OEO would be able to provide educators with a clear and reliable assessment of the capability of performance contracting to achieve the goals claimed for it by its proponents. These goals include:

-- Improving the reading and math skills of poor, under-achieving children through the use of incentive-based contracts.

-- Reducing the costs of increasing a child's achievement by certain grade levels.

-- Effecting institutional change by introducing new techniques and instructional devices into the classroom, and by developing an awareness among school officials of the need to establish educational objectives and determine whether those objectives are being met.

In addition, the experiment was designed to examine a number of related issues, such as the impact of performance contracting on school attendance and parental attitudes toward special education programs and education in general.

School Selection Process

Invitations to participate in the experiment were sent to about 200 school districts that had expressed some interest in performance contracting to the OEO, to the experiment's management support contractor (Education Turnkey Systems), or to the U.S. Office of Education. Of those 200, some 163 districts responded to the invitation, and 77 made a formal application.

To be selected, the school districts had to meet the following criteria:

-- Designate elementary and junior high schools for the experiment that met the criteria for assistance under Title I of the Elementary and Secondary Education Act.

-- Have at least 200 children each in grades 1, 2, 3, 7, 8 and 9 (100 for the experimental group and 100 for the control group).2/

-- Be able to provide data on student achievement and to provide space and personnel for the experiment.

-- Indicate that it anticipated no legal or political obstacles to mounting the experiment.

The need to include all major geographic sections of the country and to ensure representation of all major demographic subgroups of the poverty population was also considered in selecting districts. As a

---

2/ This criterion was reduced to 75 students in three cases to allow small, rural districts to participate in the experiment.

result of the screening process, 18 school districts were chosen: four serving major urban areas (Bronx, Philadelphia, Seattle, and Dallas), nine middle-sized urban systems (Anchorage, Alaska; Fresno California; Grand Rapids, Michigan; Hammond, Indiana; Hartford, Connecticut; Jacksonville, Florida; Las Vegas, Nevada; Portland, Maine; and Wichita, Kansas), and five smaller and rural systems (Athens, Georgia; McComb, Mississippi; Rockland, Maine; Selmer, Tennessee; and Taft, Texas).[3/] Their student populations included poor whites, blacks, Chicanos, Puerto Ricans, Eskimos, and Indians.

Technology Company Selection

Of the 31 technology firms responding to the OEO's request for proposals, six were selected on the basis of their corporate experience and interest in performance contracting, the types of achievement they thought they could guarantee, the qualifications of their staff, and the variety they represented in terms of their instructional approach (i.e., emphasis on hardware, incentives, or curricular software and teacher training methods). The six firms selected were: Alpha Learning Systems, Inc.; Singer/Graflex, Inc.; Westinghouse Learning Corporation; Quality Education Development, Inc.; Learning Foundations, Inc.; and Plan Education Centers, Inc. Each of the six was assigned

---

3/ The control schools for Rockland and Taft were located in nearby school districts.

to three demographically different districts among the 18. A summary of each of the firm's instructional approaches is shown in Table I.

Student Selection

The schools in each district that had the most academically deficient student bodies and which were logistically best able to accommodate the experiment were chosen to provide the experimental group; the next most deficient was chosen for the control group. Different schools were selected for the control and experimental groups, to prevent any "rub off" effects; i.e., to prevent any confounding of the data as a result of influences the performance contracting program might have on adjacent classrooms. Since the "rub off" effect might be important in its own right, however, small comparison groups also were established in each of the experimental schools. (Students in these comparison groups were also to be used as a replacement pool for students in the experimental group who might move from the district or leave the program for any other reason.) Finally, in Grand Rapids and Hartford, "special treatment" groups were identified. These included students already enrolled in special reading and math programs.

Using achievement test data supplied by the schools, the 100 students in each grade who were the farthest below grade level in reading and math were assigned to the experimental and control groups in each school. The 50 students with the next lowest scores were

Table I. Comparison of Particular Aspects of Experimental Programs[a]

Company	Average Percent of Paraprofessionals	Average Student/ Teacher Ratio	Use of Incentives	Instruction
ALPHA	Elementary: 45% Junior High: 35%	Elementary: 1/14 Junior High: 1/15	Heavy emphasis on tokens and play money which could be traded for a variety of prizes or free time	Heavy use of programmed materials and individual instruction
Singer/ Graflex	Elementary: 55% (Does not include McComb: 9%) Junior High: 32% (Does not include McComb: 6%)	Elementary: 1/20 Junior High: 1/20	Incentives used moderately and awarded infrequently	Primary reliance on programmed materials and group instruction
QED	Elementary: 50% Junior High: 50%	Elementary: 1/13 Junior High: 1/13	Heavy reliance on all types of incentives awarded randomly at least once a week	Heavy use of programmed materials with tape and cassette teaching machines
Learning Foundations	Elementary: 100% Junior High: 100%	Elementary: 1/5 Junior High: 1/6	Incentives used only moderately but awarded frequently	Used programmed materials and individual or group instruction
Plan Education Centers	Elementary: 50% Junior High: 50%	Elementary: 1/6 Junior High: 1/5	No incentives used	Used primarily teacher instruction with some tapes and cassettes
Westinghouse Learning Centers	Elementary: 80% Junior High: 80%	Elementary: 1/12 Junior High: 1/12	Heavy use of incentives in elementary grades, but few incentives used in junior high. Incentives awarded infrequently at random	Heavy reliance on programmed materials with tape and cassette teaching machines in individual and group instruction

[a]The information in this table is based on the responses of local project directors to a questionnaire.

assigned to the comparison groups. (In the case of first graders, kindergarten teachers' recommendations, readiness scores, and low-income status were usually used as the criteria for placement, since achievement scores usually were not available.)

All of the students selected initially, of course, did not participate in the experiment since some had moved from the district after school ended in June and before the experiment began in September. Replacements for students who left the experiment after the beginning of the school year were, for the most part, selected from the comparison groups.[4/]

A breakdown of the racial and ethnic composition of the control and experimental groups and of their families' per capita income is shown in Table II.

Evaluation Design

To develop an accurate gauge of performance contracting's capabilities, and to prevent "teaching to the tests," an elaborate evaluation structure was devised. Two sets of tests were used in the experiment, one for determining the private firms' pay and one for the OEO's evaluation purposes. Three different, nationally normed standardized tests, one of which was selected on a random basis for

---

4/ Indeed, the comparison groups were used so extensively as a replacement pool that their value for comparative purposes was almost completely diminished.

Table II

Characteristics of Students[a]

Site	Race or Ethnic Origin: White	Black	Spanish-Speaking	Other	Median Per Capita Income
Anchorage					
E	54%	18%	2%	26%[b]	$2,300
C	91	0	1	8	3,000
Athens					
E	37	63	0	0	1,000
C	59	41	0	0	1,250
Bronx					
E	8	42	42	8	1,400
C	2	46	50	2	1,140
Dallas					
E	0	100	0	0	700
C	0	98	2	0	570
Fresno					
E	29	11	58	2	1,070
C	43	3	53	1	1,300
Grand Rapids					
E	47	41	9	3	1,230
C	56	37	6	1	1,490
Hammond					
E	57	41	2	0	1,590
C	87	72	1	0	1,800
Hartford					
E	1	86	13	0	750
C	5	74	19	1	950
Jacksonville					
E	0	100	0	0	820
C	0	100	0	0	780
Las Vegas					
E	44	45	9	2	1,700
C	47	46	5	2	1,660

E = Experimental Group C = Control Group

[a]Based on responses to parental questionnaires for students enrolled in the experiment for the full year.

[b]Primarily Eskimo.

Table II

Characteristics of Students (Cont'd)

Site	Race or Ethnic Origin				Median Per Capita Income
	White	Black	Spanish-Speaking	Other	
McComb					
E	6%	94%	0%	0%	$ 650
C	49	51	0	0	860
Philadelphia					
E	1	96	3	0	730
C	3	92	3	3	730
Portland					
E	98	2	0	0	1,190
C	98	2	0	0	1,550
Rockland					
E	100	0	0	0	1,520
C	NA[c]	NA[c]	NA[c]	NA[c]	NA[c]
Seattle					
E	61	30	0	9[d]	1,570
C	88	7	0	5	1,900
Selmer					
E	88	12	0	0	1,390
C	92	8	0	0	1,100
Taft					
E	1	2	97	0	600
C	5	2	89	4	690
Wichita					
E	40	58	2	0	1,450
C	52	47	1	0	1,410

[c]Primarily Indian.

[d]The control students were in a different district from the experimental students. School officials in the control district refused to allow the parental questionnaire, on which these data are based, to be administered.

each class, were used for determining about 75 percent of the firms' pay, with the remainder of the pay determined by students' performance on criterion, or curriculum, referenced tests. A fourth standardized test was used only for evaluation purposes.

Both the evaluation and the payments standardized tests were chosen to:[5/]

-- Use norms that were based on a relatively recent sample having a reasonably large number of students representative of the national population.

-- Be based on a fairly recent survey of what is taught throughout the country in reading and math.

-- Display a high degree of reliability.

-- Have very clear and simple directions for administration.

It was felt by the OEO that standardized tests would provide an equitable and objective measure of the success of performance contracting, since success on such tests is strongly related to general success in school. Further, while the contractors were free to determine <u>how</u> they would attain certain objectives, the decision as to what the objectives would be was not theirs to make. Their contractual agreement to be judged on the basis of the standardized tests was an indication of their belief in the validity of the tests. Indeed, they were asked to suggest appropriate tests for the evaluation, and most of those used were the ones they suggested.

5/ A very complete description of the tests used and a discussion of the issues involved in the whole testing question will be included in OEO pamphlet 3400-6.

The payments tests, which were administered only to the experimental group, were given within the first 10 days and the last 15 days of school at each site. The evaluation tests, which were given to the experimental, control, comparison, and special treatment groups, also were administered at the very beginning and very end of the school year. And, to prevent the possibility of introducing a "practice effect," the evaluation tests were administered to the experimental group before the payments tests. While both the evaluation and payments tests were primarily concerned with achievement in reading and math, the evaluation tests also measured students' performance in science, social studies, spelling, and language skills.

Several safeguards were built into the evaluation structure to prevent "teaching to the tests." The companies did not know, and were threatened with penalties for attempting to learn, which form of the standardized tests was used. Company personnel were not involved in administering or scoring the tests. To prevent any inadvertent use of material containing test items, the management support contractor conducted curriculum audits on a spot basis. In addition, to determine whether initial results were retained, retention tests were administered on a selective basis during the current school year.

Some 25 percent of the contractors' pay was based on the results of interim performance objective tests (IPOs), which were given five times during the year to assess the students' mastery of the specific

curricular materials to which they had been exposed. The IPOs were added to the payments structure because it was felt they would offer a useful supplement to the standardized tests for payment purposes and that they might add to the overall evaluation. It was intended that the firms submit a pool of potential IPO test items to the evaluation contractor, Battelle Memorial Institute, and that Battelle randomly select one-third of those items for the actual testing.

In practice, however, these intentions could not be carried out. First, the firms' heavy reliance on individualized instruction, and hence the need for an unmanageable number of different tests, made the requirement of tripling the number of test items unworkable. Second, the firms' freedom to change their curricula during the course of the school year made the requirement of submitting test items in advance unrealistic. As a result, Battelle did not review IPO test items before they were administered. Consequently, it seems that some of the tests were too easy; in one site in one grade/subject combination, less than 1 percent of the children failed to answer at least 75 percent of the questions correctly. In addition, it would appear that not all the tests were relevant measures of what the contractors had taught. In a few instances, Battelle initially refused to certify the tests, but since Battelle's review took place after the tests were administered, nothing could be done to correct the problem.

Thus, the IPOs appear to have been virtually useless for evaluation purposes and to have had only questionable value for payment purposes.

RESULTS

The single most important question for all concerned with the experiment is: Was performance contracting more successful than traditional classroom methods in improving the reading and math skills of poor children? The answer, as shown in Table III, is: No.

The analysis summarized in the table is based on the average (or mean) grade level gains of all students in the experimental and control groups who took both the pre- and post-experiment evaluation tests.[6/] The right-hand column of the table demonstrates that the difference in gains was remarkably small in all 10 of the grade/subject combinations for which this analysis is appropriate. In half of the 10 cases, there was no difference at all between the gains of the experimental and control groups. In four of the cases, there was a difference of only one-tenth of a grade level, and in only one case was there a difference of as much as two-tenths of a grade level. These overall differences are so slight that we can conclude that performance contracting was no more effective in either reading or math than the traditional classroom methods of instruction.

Table III also indicates that the performance of students in the experimental group does not appear disappointing just because students

---

6/ The number of children who took both the pre- and post- tests represents only about two-thirds of those who were enrolled initially. Many children moved away or dropped out of school during the year and, while they usually were replaced by others, the replacements often entered the program too late for their performance to be meaningful for analytical purposes. Other children were absent when either or both of the evaluation tests were administered.

Table III

Mean Gains of Experimental and Control Students
Across All Sites

Reading

	Experimental Gain	Control Gain	Difference
Grade 1	NA	NA	NA
2	.4	.5	-.1
3	.3	.2	+.1
7	.4	.3	+.1
8	.9	1.0	-.1
9	.8	.8	--

Math

	Experimental Gain	Control Gain	Difference
Grade 1	NA	NA	NA
2	.5	.5	--
3	.4	.4	--
7	.6	.6	--
8	.8	1.0	-.2
9	.8	.8	--

NA: A readiness test, rather than an achievement test, was used as the first grade pretest. There is no grade equivalent for the readiness test.

in the control group did unexpectedly well. In fact, neither group did well. In only two of the 20 possible cases was the mean gain of either the control or experimental students as much as one grade level.

Table IV looks at the results from a slightly different perspective, showing the mean grade levels of children in the experimental group at the beginning and end of the experiment. From this table, it can be seen that performance contracting was not successful in meeting its original goal of bringing under-achieving students' performance up to grade level. In all cases, the average achievement level of children in the experimental group was well below the norm for their grade and in all cases, in terms of grade equivalents, the average slipped even further behind during the year.

Thus, it is fairly clear that regardless of the perspective taken, performance contracting was not responsible for any significant improvement on an overall basis. The next logical question then, is: Do the overall results mask individual success stories among certain types of students or students in certain sites?

One way to analyze whether performance contracting was particularly successful among certain types of students is to examine its impact on the scores of children at various points on the distribution; that is, to look at its effect on the score of the child who is at the 20th, 40th, 50th, 60th, and 80th percentiles. Table V, by way of example, shows the results of this analysis by comparing the pretest and post-test levels of the third grade students in reading at the various percentile rankings. From this table, it can be seen that the differences in levels

Table IV

Status of Experimental Students
Before and After Performance Contracting

Reading

	Starting Position[a]	Ending Position	Relation to Grade Level at End
Grade 1	NA	1.0	- .9
2	1.5	1.9	-1.0
3	2.2	2.5	-1.4
7	4.5	4.9	-3.0
8	4.8	5.7	-3.2
9	5.6	6.4	-3.5

Math

	Starting Position[a]	Ending Position	Relation to Grade Level at End
Grade 1	NA	1.3	- .6
2	1.4	1.9	-1.0
3	2.2	2.6	-1.3
7	4.7	5.3	-2.6
8	5.4	6.2	-2.7
9	6.0	6.8	-3.1

[a] Pretest grade equivalent rating not available for first grade students.

Table V

Evaluation Test Results (GEQ)

3rd Grade Reading

PERCENTILE	EXPERIMENTAL			CONTROL		
	Pre	Post	Difference	Pre	Post	Difference
20	1.7	2.1	.4	1.7	2.2	.5
40	1.9	2.4	.5	2.1	2.5	.4
50	2.0	2.5	.5	2.2	2.8	.6
60	2.2	2.7	.5	2.4	3.1	.7
80	2.4	3.2	.8	2.8	3.6	.8

are very similar for children at all points in the distribution. Although the other 11 grade/subject combinations are not presented here, they have been examined and, again, the results are similar. No significantly different impacts were discovered among children at different points on the distribution. In other words, there is no evidence that performance contracting had differential results for the lowest or highest achieving students in the sample.

Table VI attempts to show whether a number of dramatically "good" sites offset a number of dramatically "bad" sites to produce the overall neutral effect. The data for this table were generated by comparing the differences in mean gains for experimental and control groups at each site. These comparisons of individual site results are considerably less reliable than overall conclusions, because testing conditions were less than ideal at some sites; at others, control group students seem to have performed inexplicably poorly or well; and at others, the pre-test scores of the experimental and control students were not pertectly matched. These problems do, for the most part, offset each other in the overall comparisons. Nevertheless, a summary of individual site effects can give a crude estimate of whether many successes or many failures were masked by the overall results.

Again, this does not appear to be the case. While there were a few apparent successes or failures among the sites, in 80 percent of the cases, there was no evidence of significant differences in the gains of

Table VI

Summary of Significant Results at Individual Sites[a]

Reading

	Significant Gains	Significant Losses	No Significant Difference
Grade 1	5	4	8
2	-	-	18
3	1	-	17
7	-	1	17
8	3	2	13
9	6	1	10
Total	15	8	83

Math

	Significant Gains	Significant Losses	No Significant Difference
Grade 1	4	4	9
2	-	-	18
3	2	1	15
7	-	2	16
8	1	2	15
9	4	2	11
Total	11	11	84

[a] A significant gain or loss is defined as being a relative improvement of one-half grade level equivalent or more.

the experimental and control groups.

Thus, despite all the uncertainties that inevitably surround anything involving the testing of human beings, the results from the performance contracting experiment point with remarkable consistency to the conclusion that there were no significant differences in the achievement gains of the experimental and control groups. Not only did both groups do equally poorly in terms of overall averages, but also these averages were very nearly the same in each grade, in each subject, for the best and worst students in the sample, and, with few exceptions, in each site. Indeed, the most interesting aspect of these conclusions is their very consistency. Thus, the evidence does not indicate that performance contracting will bring about any great improvement in the educational status of disadvantaged children.

7/ The analyses presented in this section are, of course, the result of rather straightforward comparisons. Because experimental and control groups were not randomly assigned, and differ somewhat in their characteristics, more complicated multivariate analyses were initially thought to be appropriate. Many different analyses have been performed and measurement error as well as biases introduced by the mismatch of the two groups were examined. Our judgment is that the simple comparisons reported here are as unbiased as any of the more complex approaches. In any case, none of the analyses performed indicated different overall results, although in some cases they altered the relative "success" or "failure" of specific site/grade/subject combinations. An extensive discussion of these analyses will be included in the forthcoming OEO pamphlet 3400-6.

CONTRACTUAL PROCEDURES

As noted earlier, the performance contract itself was the crux of this new education concept. Under a performance contract, unlike common cost-reimbursable contracts, payments to firms are not based upon actual costs. Instead, earnings are determined by the performance of the children whom they instruct.

All of the contracts in this experiment included identical general provisions, including statements of work, responsibilities of the private firms and schools, and procedures for testing and student selection and attendance. Each of the contracts also specified that up to 75 percent of the payments would be based on the results of standardized tests and up to 25 percent on the interim, criterion referenced tests. In addition, the maximum that a firm could earn in total was based on a figure of about $200 per student per subject. The $200 figure was chosen by the Office of Economic Opportunity to proximate, roughly, current public school per student expenditures on reading and math instruction and to set a budget constraint that would be affordable by public schools should they decide to replicate the experimental programs. During the contract negotiations with the firms, the $200 figure was adjusted for each contract to reflect local conditions, such as teacher salary scales and cost of living indices, so that in actuality, the base figure for different sites ranged from $185 to $240.[8/]

8/ The base for Alpha was $165, since in Alpha's programs certified teachers were employees of the participating districts, and their salaries were not part of Alpha's costs. Paraprofessionals were on Alpha's payroll.

As noted earlier, up to 25 percent of the total contract price could be earned on the basis of students' performance on the IPOs, and the remainder on the basis of their performance on the standardized test. The determination of whether the contractor had earned the 25 percent was relatively simple: The firm received one-fifth of that amount for each child each time the child passed one of the five IPO tests that were given during the year. The determination of whether the contractor had earned the remaining 75 percent, or any portion of that amount, was more complex. Two factors were taken into account in making that determination:

-- How many children had improved in reading and math by a certain level set in advance. When the private firms submitted their bids, they indicated a minimum level of improvement they would guarantee in each subject in each grade. This minimum guarantee, which had to be achieved before the contractor was eligible to receive any payment for a particular student, ranged from a half a grade level to one and a half grade levels.9/

9/ These minimum guarantee levels should be viewed in light of the fact that most children in the experiment were at least one grade level below norm before the experiment began, with the decrement generally increasing among the higher grades. As table IV shows, the mean decrement in reading among ninth graders was three grade levels, meaning that the average student entering the ninth grade at the beginning of the experiment was reading at between the fifth and sixth grade level. The improvement that normally could be expected among students with similar achievement records is less than a grade level per year. The private firms, then, typically had to do better than this to receive any payment at all, and much better than this to earn a profit.

contractors also had specified amounts, ranging from $46.25 to $101.00 per child per subject, that they would receive for all students who improved by at least the minimum level that had been guaranteed, as shown in Table VII .

-- Improvement beyond the minimum guarantee level. In addition, the contractors set the dollar amount they would receive for each tenth of a grade level each child advanced above the minimum guarantee level. The amounts the contractors received for those incremental increases ranged from $5.36 to $20.00 per one-tenth of a grade level improvement.

The incentive scale was structured so that the contractors' pay was based on the performance of each individual child, rather that class or site averages. If one child achieved the minimum improvement, the contractor would be paid for that child. If the next child did not improve, the contractor would not be paid for that child. No ceiling was set on the amount a contractor could earn for an individual child's improvement. Rather, a ceiling was set on the maximum a contractor could earn at any one site.

By mid-February, it became apparent that some changes in the original contracts would have to be negotiated to account for unanticipated problems facing the private firms. For example, the original terms specified that a definite number of students would be present for definite periods of instruction. Teacher strikes, absenteeism, bad weather, student drop-outs, and other factors made it impossible for school districts to fulfill those guarantees. Adjustments for these factors are presently being negotiated.

Table VII

Summary of Contractor Incentive Scales

Contractor	Minimum Guaranteed Gain (Grade Equivalent on Standardized Tests	Price for Minimum Gain	Price per 0.1 Above Minimum Gain
Alpha	0.8 (Gr. 1-3)[a]	$56.25	$6.25[b]
	1.0 (Gr. 7-9)	75.00	5.36[b]
Learning Foundations	1.0 (Gr. 1-3)	101.00	8.77
	1.1 (Gr. 7-9)	81.00	8.25
Plan	0.5 (Gr. 1-Math)	50.00	20.00
	0.5 (Gr. 1-Read)	46.25	9.25
	1.0 (Gr. 2,3-Math)	50.00	20.00
	1.0 (Gr. 2,3-Read)	46.25	9.25
	1.0 (Gr. 7,9-Math)	50.00	10.00
	1.0 (Gr. 7,9-Read)	55.00	5.50
QED	1.0 (Gr. 1-3)	72.50	8.50
	1.5 (Gr. 7-9)	82.50	15.00
Singer	0.5 (Gr. 1,2)	82.50	8.25
	1.0 (Gr. 3,7-9)	82.50	7.17
Westinghouse	1.0	75.00	10.70

NOTES: Prices shown are representative of all school districts for each contractor; if its prices varied by district, the lowest price is shown. Guarantee schedules for each contractor did not vary by district except where noted.

[a]0.5 minimum guarantee in Taft, Texas.

[b]The actual price per 0.1 above the minimum was varied at different points in the scale. Figure shown is the average.

During the negotiation process, it has become apparent that the terms of the initial contracts allowed too much room for difference in interpretation, for example, and that the roles of the various experiment participants were not spelled out clearly enough.

It has also become clear that more attention needs to be paid to the incentive structure incorporated into the contracts. The structure of the Office of Economic Opportunity's contracts, outlined above, seemed entirely reasonable -- pay nothing unless a student reaches a significant minimum gain level and then reward the contractor for performance above this point. Yet this structure implies some rather questionable assumptions about educational objectives. Specifically, it implies that we are indifferent as to whether a student gains .1 year or .9 of a year, as long as he remains below the minimum guaranteed gain, and that we value equally a one year gain for a student who is one year behind and for a student who is four years behind. In addition, depending on the specific contract terms, in many cases, it implies that we are essentially indifferent as to whether all the students gain 1½ years, whether half the students gain no year and half gain two years, or whether half gain less than a year and half gain three years. These may well reflect reasonable educational objectives -- but we doubt it. Yet the structure was adopted by the contractors and by many other

school systems and has not, to our knowledge, been seriously questioned by anyone. And we also doubt that many school systems have given much attention to thinking about their objectives in these terms. While measureable skills such as reading and math clearly constitute only a part of the objectives of any school system, we feel much more attention should be given to specifying such objectives -- and measuring performance against them -- on a systemwide basis.

SUMMARY AND CONCLUSIONS

In considering the implications of the results presented here, it is important to reiterate what was being tested in the experiment:

-- The capabilities of a representative group of private education firms using existing instructional materials and technologies and working under specific kind of performance-based contract.

-- A concept that proponents hoped would be more effective than traditional classroom methods in improving the reading and math skills of poor, under-achieving children.

The results of the experiment clearly indicate that the firms operating under performance contracts did not perform significantly better than the more traditional school systems. Indeed, both control and experimental students did equally poorly in terms of achievement gains, and this result was remarkably consistent across sites and among children with different degrees of initial capability. On the basis of these findings it is clear that there is no evidence to support a massive move to utilize performance contracting for remedial education in the nation's schools. School districts should be skeptical of extravagant claims for the concept.

At the same time, the results should not be interpreted as a blanket finding that educational services and materials should not be purchased under performance-based contracts or that private firms cannot

provide valuable educational services. Surely performanced based contracts are in some cases a better way to purchase some educational services than the methods currently being used. Surely private firms should continue to play an important role in developing and marketing new educational materials. The results simply say that an uncritical rush to embrace these concepts is unwarranted at this time.

Some of the benefits of this experiment will not be known for sometime, and indeed cannot be precisely pinpointed. The experiment has provoked or added to useful debates on the current use of standardized tests for measuring student performance, on means of introducing change into the educational system, and in general on the subject of accountability. It has raised the possibility that other performers besides schools may sometimes be appropriate providers of education. And hopefully, it will lead to a heightened awareness of the importance of specifying educational goals and measuring progress toward those goals, a process that all too frequently has not been undertaken by school districts.

But surely the clearest conclusion drawn from the experiment is that we still have no solutions to the specific problem of teaching disadvantaged youngsters basic math and reading skills. Thus while we judge this experiment to be a success in terms of the information it can offer about the capabilities of performance contractors, it is clearly another failure in our search for means of helping poor and disadvantaged youngsters to develop the skills they need to lift themselves out of poverty. The search for solutions to these problems must continue.